nature genetics

Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis

To elucidate the genetic architecture of amyotrophic lateral sclerosis (ALS) and find associated loci, we assembled a custom imputation reference panel from whole-genome-sequenced patients with ALS and matched controls (*n* = 1,861). Through imputation and mixed-model association analysis in 12,577 cases and 23,475 controls, combined with 2,579 cases and 2,767 controls in an independent replication cohort, we fine-mapped a new risk locus on chromosome 21 and identified *C21orf2* as a gene associated with ALS risk. In addition, we identified *MOBP* and *SCFD1* as new associated risk loci. We established evidence of ALS being a complex genetic trait with a polygenic architecture. Furthermore, we estimated the SNP-based heritability at 8.5%, with a distinct and important role for low-frequency variants (frequency 1–10%). This study motivates the interrogation of larger samples with full genome coverage to identify rare causal variants that underpin ALS risk.

ALS is a fatal neurodegenerative disease that affects 1 in 400 people, with death occurring within 3 to 5 years of the onset of symptoms[1]. Twin-based studies estimate heritability to be around 65%, and 5–10% of patients with ALS have a positive family history[1,2]. Both of these features are indicative of an important genetic component in ALS etiology. Following initial discovery of a risk-associated *C9orf72* locus in ALS genome-wide association studies (GWAS)[3–5], identification of a pathogenic hexanucleotide-repeat expansion in this locus revolutionized the field of ALS genetics and biology[6,7]. The majority of ALS heritability, however, remains unexplained, and only two additional risk loci have since been identified robustly[3,8].

To discover new genetic risk loci and elucidate the genetic architecture of ALS, we genotyped 7,763 new cases and 4,669 controls and additionally collected genotype data from published GWAS of ALS. In total, we analyzed 14,791 cases and 26,898 controls from 41 cohorts (**Supplementary Table 1** and **Supplementary Note**). We combined these cohorts on the basis of genotyping platform and nationality to form 27 case–control strata. In total, 12,577 cases and 23,475 controls passed quality control (Online Methods and **Supplementary Tables 2–5**).

For imputation purposes, we obtained high-coverage (~43.7×) whole-genome sequencing data from 1,246 patients with ALS and 615 controls from the Netherlands (Online Methods and **Supplementary Fig. 1**). After quality control, we constructed a reference panel including 18,741,510 single-nucleotide variants (SNVs). Imputing this custom reference panel into Dutch ALS cases considerably increased the imputation accuracy for low-frequency variants (minor allele frequency (MAF) = 0.5–10%) in comparison to commonly used reference panels from 1000 Genomes Project Phase 1 (ref. 9) and Genome of the Netherlands[10] (**Fig. 1a**). Improvement was also observed when imputing into ALS cases from the UK (**Fig. 1b**). To benefit from the global diversity of haplotypes, the custom and 1000 Genomes Project panels were combined, which further improved imputation. Given these results, we used the merged reference panel to impute all strata in our study.

In total, we imputed 8,697,640 variants passing quality control into the 27 strata and tested the strata separately for association with ALS risk by logistic regression. We then included the results in an inverse-variance-weighted, fixed-effects meta-analysis, which identified four loci associated at genome-wide significance ($P < 5 \times 10^{-8}$) (**Fig. 2a**). The previously reported *C9orf72* (rs3849943)[3–5,8], *UNC13A* (rs12608932)[3,5] and *SARM1* (rs35714695)[8] loci all reached genome-wide significance, as did a new association for a nonsynonymous variant in *C21orf2* (rs75087725, $P = 8.7 \times 10^{-11}$; **Supplementary Tables 6–10**). This variant was present on only 10 haplotypes in the 1000 Genomes Project reference panel (MAF = 1.3%), whereas it was present on 62 haplotypes in our custom reference panel (MAF = 1.7%). As a result, more strata passed quality control for this variant by passing the allele frequency threshold of 1% (**Supplementary Table 11**). This result demonstrates the benefit of the merged reference panel with ALS-specific content, which improved imputation and resulted in the identification of a genome-wide significant association.

Linear mixed models (LMMs) can improve power while controlling for sample structure[11], which would be particularly important in our study that included a large number of imperfectly balanced strata. Even though LMM analysis for ascertained case–control data potentially results in a small loss of power in comparison to meta-analysis[11], we judged the advantage of combining all strata while controlling the false positive rate to be more important than this potential loss and therefore jointly analyzed all strata in an LMM to identify additional risk loci. There was no overall inflation of the LMM test statistics in comparison to the meta-analysis test statistics (**Supplementary Fig. 2**). We observed modest inflation of test statistics in the quantile–quantile plot ($\lambda_{GC} = 1.12$, $\lambda_{1,000} = 1.01$; **Supplementary Fig. 3**). LD score regression yielded an intercept of 1.10 (standard error

A full list of authors and affiliations appears at the end of the paper.

Received 7 January; accepted 20 June; published online 25 July 2016; doi:10.1038/ng.3622

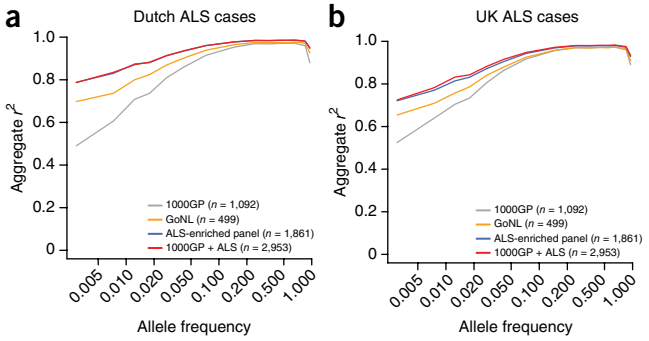

Figure 1 Comparison of imputation accuracy. (**a**,**b**) Aggregate r^2 values between imputed and sequenced genotypes on chromosome 20 are shown when using different reference panels for imputation. Allele frequencies were calculated from the Dutch samples included in the Genome of the Netherlands (GoNL) cohort. The highest imputation accuracy was achieved when imputing from the merged custom and 1000 Genomes Project (1000GP) panel. The difference in accuracy was most pronounced for low-frequency alleles (frequency 0.5–10%) in ALS cases from both the Netherlands (**a**) and the UK (**b**).

of 7.8×10^{-3}). Although an LD score regression intercept higher than 1.0 can indicate the presence of residual population stratification, which is fully corrected for in an LMM, this can also reflect a distinct genetic architecture where most causal variants are rare or a noninfinitesimal architecture[12]. The LMM identified all four genome-wide-significant associations from the meta-analysis. Furthermore, three additional loci—*MOBP* at 3p22.1 (rs616147), *SCFD1* at 14q12 (rs10139154) and a long noncoding RNA at 8p23.2 (rs7813314)—were associated at genome-wide significance (**Fig. 2b**, **Table 1** and **Supplementary Tables 12–14**). SNPs in the *MOBP* locus have been reported to be associated in a GWAS on progressive supranuclear palsy (PSP)[13] and to act as a modifier for survival in frontotemporal dementia (FTD)[14]. The putative pleiotropic effects of variants in this locus suggest that ALS, FTD and PSP share a neurodegenerative pathway. We also found that rs74654358 at 12q14.2 in the *TBK1* gene approximated genome-wide significance (MAF = 4.9%, odds ratio (OR) = 1.21 for the A allele, $P = 6.6 \times 10^{-8}$). This gene was recently identified as an ALS risk gene through exome sequencing[15,16].

In the replication phase, we genotyped the newly discovered associated SNPs in nine independent replication cohorts, totaling 2,579 cases and 2,767 controls. In these cohorts, we replicated the signals for the *C21orf2*, *MOBP* and *SCFD1* loci, with lower *P* values in the combined analysis than in the discovery phase (combined

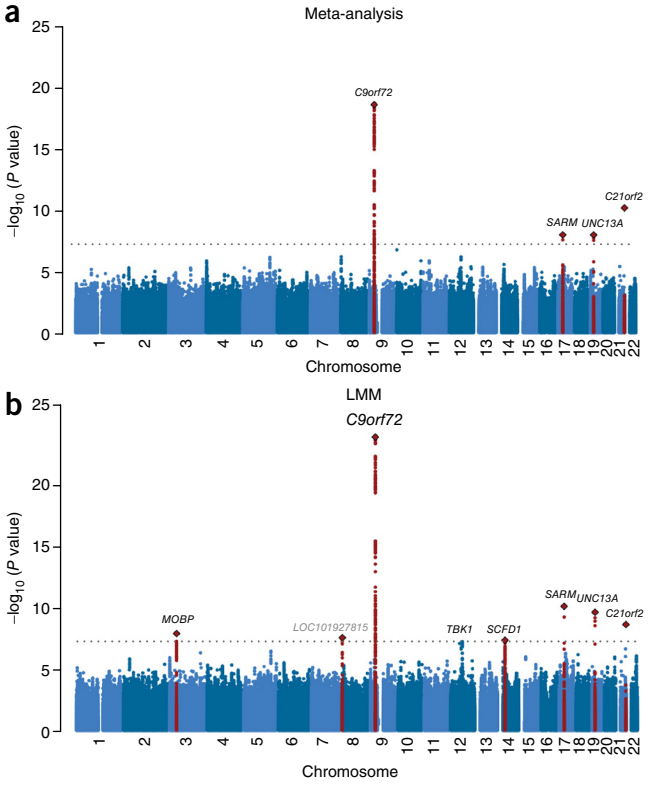

Figure 2 Meta-analysis and LMM associations. (**a**) Manhattan plot for the meta-analysis results. This approach yielded four genome-wide-significant associations. The associated SNP in *C21orf2* is a nonsynonymous variant not found to be associated in previous GWAS. (**b**) Manhattan plot for the LMM results. This analysis yielded three loci in addition to those identified by meta-analysis with associations that reached genome-wide significance (*MOBP*, *LOC101927815* and *SCFD1*). The association for SNPs in the previously identified ALS risk gene *TBK1* approached genome-wide significance ($P = 6.6 \times 10^{-8}$). As the *C21orf2* SNP was removed from a Swedish stratum because of MAF <1%, this SNP was tested separately, but it is presented here together with all SNPs with MAF >1% in all strata. *LOC101927815* is shown in gray because the association for this locus could not be replicated. Loci are labeled by the name of the nearest gene. The dotted lines correspond to the significance threshold of $P = 5 \times 10^{-8}$.

P value = 3.08×10^{-10}, 4.19×10^{-10} and 3.45×10^{-8} for rs75087725, rs616147 and rs10139154, respectively; **Table 1** and **Supplementary Fig. 4**)[17]. The combined signal for rs7813314 was less significant because the effects for the discovery and replication phases were in opposite directions, indicating non-replication. Although replication yielded an effect estimate for rs10139154 similar to that obtained in the discovery phase, this effect was not statistically significant ($P = 0.09$) in the replication phase alone. This lack of significance reflects the limited sample size of our replication phase, a feature that is inherent to studies of ALS because of its low prevalence. Even larger sample sizes are warranted to replicate this signal robustly.

There was no evidence of residual association in each locus after conditioning on the top SNP, indicating that all the risk loci are independent signals. Apart from the *C9orf72*, *UNC13A* and *SARM1* loci, we found no evidence of associations previously described in smaller GWAS (**Supplementary Table 15**).

The association of the low-frequency nonsynonymous SNP in *C21orf2* suggested that this gene could be directly involved in ALS risk. Indeed, we found no evidence that linkage disequilibrium (LD) between this SNP and sequenced variants beyond the boundaries of *C21orf2* explained the association of this locus (**Supplementary Fig. 5**). In addition, we investigated the burden of rare coding mutations in *C21orf2* in a set of whole-genome-sequenced cases ($n = 2{,}562$) and controls ($n = 1{,}138$). After quality control, these variants were tested for association using pooled association tests for rare variants and applying correction for population structure (tests T5 and T1 for alleles with 5% and 1% frequency, respectively; **Supplementary Note**). This approach demonstrated an excess of nonsynonymous and loss-of-function mutations in *C21orf2* among ALS cases that persisted after conditioning on rs75087725 ($P_{T5} = 9.2 \times 10^{-5}$, $P_{T1} = 0.01$; **Supplementary Fig. 6**), further supporting the notion that *C21orf2* contributes to ALS risk.

In an effort to fine-map the other loci to pinpoint susceptibility genes, we searched for SNPs in these loci with *cis* expression quantitative

Table 1 Discovery and replication of new genome-wide significant loci

	Discovery					Replication				Combined	
SNP	MAF_{cases}	$MAF_{controls}$	OR	P_{meta}	P_{LMM}	MAF_{cases}	$MAF_{controls}$	OR	P	$P_{combined}$	I^2
rs75087725	0.02	0.01	1.45	8.65×10^{-11}	2.65×10^{-9}	0.02	0.01	1.65	3.89×10^{-3}	3.08×10^{-10}	0.00*
rs616147	0.30	0.28	1.10	4.14×10^{-5}	1.43×10^{-8}	0.31	0.28	1.13	2.35×10^{-3}	4.19×10^{-10}	0.00*
rs10139154	0.34	0.31	1.09	1.92×10^{-5}	4.95×10^{-8}	0.33	0.31	1.06	9.55×10^{-2}	3.45×10^{-8}	0.05*
rs7813314	0.09	0.10	0.87	7.46×10^{-7}	3.14×10^{-8}	0.12	0.10	1.17	7.75×10^{-3}	1.05×10^{-5}	0.80**

Genome-wide-significant loci from the discovery phase including 12,557 cases and 23,475 controls were directly genotyped and tested for association in the replication phase including 2,579 cases and 2,767 controls. The three top associated SNPs in the *MOBP* (rs616147), *SCFD1* (rs10139154) and *C21orf2* (rs75087725) loci replicated with associations in the same direction as in the discovery phase and an association in the combined analysis that exceeded that in the discovery phase. Cochrane's Q test, $^*P > 0.1$, $^{**}P = 4.0 \times 10^{-6}$. MAF, minor allele frequency; OR, odds ratio, P_{meta}, meta-analysis P value; P_{LMM}, linear mixed-model P value; $P_{combined}$, P value from meta-analysis of the associations in the discovery and replication phase.

trait locus (*cis*-eQTL) effects observed in brain and other tissues (**Supplementary Table 16** and **Supplementary Note**)[18]. We found overlap with previously identified brain *cis*-eQTLs for five regions (**Supplementary Fig. 7**, **Supplementary Table 17** and **Supplementary Data Set**). In the *C9orf72* locus, we found that proxies of rs3849943 (LD r^2 = 0.21–0.56) only had a brain *cis*-eQTL effect on *C9orf72* (minimal $P = 5.27 \times 10^{-7}$), which harbors the hexanucleotide-repeat expansion that drives this GWAS signal. Additionally, we found that rs12608932 and its proxies in the *UNC13A* locus had an exon-level *cis*-eQTL effect on *KCNN1* in frontal cortex ($P = 1.15 \times 10^{-3}$)[19]. Another overlap was observed in the *SARM1* locus where rs35714695 and its proxies had the strongest exon-level *cis*-eQTL effect on *POLDIP2* in multiple brain tissues ($P = 2.32 \times 10^{-3}$). In the *SCFD1* locus, rs10139154 and its proxies had a *cis*-eQTL effect on *SCFD1* in cerebellar tissue ($P = 7.71 \times 10^{-4}$). For the *MOBP* locus, rs1768208 and its proxies had a *cis*-eQTL effect on *RPSA* ($P = 7.71 \times 10^{-4}$).

To describe the genetic architecture of ALS, we generated polygenic scores, which can be used to predict phenotypes for traits with a polygenic architecture[20]. We calculated SNP effects using an LMM in 18 of the 27 strata and subsequently assessed predictive ability in the other 9 independent strata. This analysis showed that a significant albeit modest proportion of the phenotypic variance could be explained by all SNPs (Nagelkerke $r^2 = 0.44\%$, $r^2 = 0.15\%$ on the liability scale, $P = 2.7 \times 10^{-10}$; **Supplementary Fig. 8**). This finding adds to the existing evidence that ALS is a complex genetic trait with a polygenic architecture. To further quantify the contribution of common SNPs to ALS risk, we estimated SNP-based heritability using three approaches, all assuming a population baseline risk of 0.25% (ref. 21). GCTA-REML estimated the SNP-based heritability at 8.5% (s.e.m. = 0.5%). Haseman–Elston regression yielded a very similar estimate of 7.9%, and LD score regression estimated the SNP-based heritability at 8.2% (s.e.m. = 0.5%). The heritability estimates for each chromosome were significantly correlated with chromosome length ($r^2 = 0.46$, $P = 4.9 \times 10^{-4}$; **Fig. 3a**), again indicative of a polygenic architecture in ALS.

We found that the genome-wide-significant loci only explained 0.2% of heritability, and the bulk of the heritability (8.3%, s.e.m. = 0.3%) was thus captured by SNPs with associations below genome-wide significance. This finding implies that many genetic risk variants have yet to be discovered. Understanding where these unidentified risk variants remain across the allele frequency spectrum will inform the design of future studies to identify these variants. We therefore estimated heritability partitioned by MAF. Furthermore, we contrasted these results with those for common polygenic traits studied in GWAS such as schizophrenia. We observed a clear trend indicating that most variance is explained by low-frequency SNPs (**Fig. 3b**). Exclusion of the *C9orf72* locus, which harbors the rare pathogenic repeat expansion, and the other genome-wide-significant loci did not affect this trend (**Supplementary Fig. 9**). This architecture is different from that expected for common polygenic traits and reflects a polygenic rare variant architecture observed in simulations[22].

To gain better insight into the biological pathways that explain the associated loci found in this study, we looked for enriched pathways using DEPICT[23]. This analysis identified SNAP receptor (SNARE) activity as the only enriched category (false discovery rate (FDR) < 0.05; **Supplementary Fig. 10**). SNARE complexes have a central role in neurotransmitter release and synaptic function[24], which are both perturbed in ALS[25].

Although the biological role of C21orf2, a conserved leucine-rich-repeat protein, remains poorly characterized, this protein is part of the ciliome and is required for the formation and/or maintenance of primary cilia[26]. Defects in primary cilia are associated with various neurological disorders, and cilia numbers are decreased in mice expressing the Gly93Ala mutant of human SOD1, a well-characterized ALS model[27]. C21orf2 has also been localized to mitochondria in immune cells[28] and is part of the interactome of the protein product of *NEK1*, which has previously been associated with ALS[15]. Both proteins seem to be involved in DNA repair mechanisms[29]. Although future studies are needed to dissect the function of *C21orf2* in ALS pathophysiology, we speculate that defects in *C21orf2* may lead to primary cilium and/or mitochondrial dysfunction or inefficient DNA repair and thereby result in adult-onset disease. The other associated loci will require more extensive studies to fine-map causal variants. *SARM1* has been suggested to be a susceptibility gene for ALS, mainly because of its role in Wallerian degeneration and its interaction with *UNC13A*[8,30]. Although these are indeed interesting observations, the brain *cis*-eQTL effect for SNPs in this locus on *POLDIP2* suggests

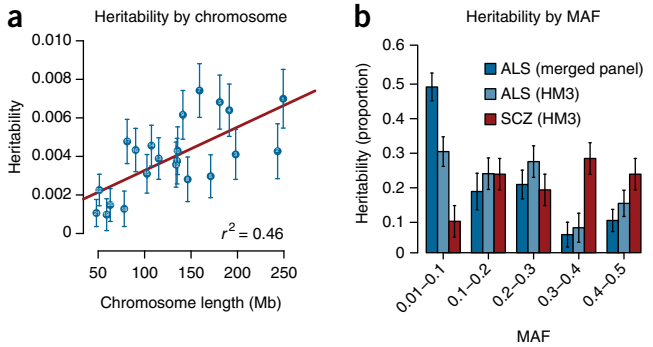

Figure 3 Partitioned heritability. (**a**) Heritability estimates for each chromosome were significantly correlated with chromosome length ($P = 4.9 \times 10^{-4}$). (**b**) For ALS, there was a clear trend where more heritability was explained by the low-frequency alleles. This effect was still observed when, for a fair comparison between ALS and a previous study partitioning heritability for schizophrenia (SCZ) using identical methods[22], SNPs present in HapMap 3 (HM3) were included. Error bars correspond to standard errors.

that *POLDIP2* and not *SARM1* could in fact be the causal gene in this locus. Similarly, *KCNN1*, which encodes a neuronal potassium channel involved in neuronal excitability, could be the causal gene either through a direct eQTL effect or rare variants in LD with the associated SNP in *UNC13A*.

In conclusion, we have identified a key role for rare variation in ALS and discovered SNPs in new complex loci. Our study therefore informs future study design in ALS genetics, promoting the combination of larger sample sizes, full genome coverage and targeted genome editing experiments, leveraged together to fine-map new loci, identify rare causal variants and thereby elucidate the biology of ALS.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The GWAS summary statistics and sequenced variants are publicly available through the Project MinE data browser at http://databrowser.projectmine.com/.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The work of the contributing groups was supported by various grants from governmental and charitable bodies. Details are provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

A.V., N.T., K.L., B.R., K.V., M.R.-G., B.K., J.Z., L.L., L.D.G., S.M., F.S., V.M., M.d.C., S. Pinto, J.S.M., R.R.-G., M.P., S. Chandran, S. Colville, R.S., K.E.M., P.J.S., J.H., R.W.O., A. Pittman, K.S., P.F., A. Malaspina, S.T., S. Petri, S. Abdulla, C.D., M.S., T. Meyer, R.A.O., K.A.S., M.W.-P., C.L.-H., V.M.V.D., J.Q.T., L.E., L. McCluskey, A.N.B., Y.P., T. Meitinger, P.L., M.R.-B., C.R.A., C. Maurel, G. Bensimon, B.L., A.B., C.A.M.P., S.S.-D., A.D., N.W.W., L.T., W.L., A.F., M.R., S. Cichon, M.M.N., P.A., C. Tzourio, J.-F.D., A.G.U., F.R., K.E., A.H., C. Curtis, H.M.B., A.J.v.d.K., M.d.V., A.G., M.W., C.E.S., B.N.S., O.P., C. Cereda, R.D.B., G.P.C., S.D'A., C.B., G.S., L. Mazzini, V.P., C.G., C. Tiloca, A.R., A. Calvo, C. Moglia, M.B., S. Arcuti, R.C., C.Z., C.L., S. Penco, N.R., A. Padovani, M.F., B.M., R.J.S., PARALS Registry, SLALOM Group, SLAP Registry, FALS Sequencing Consortium, SLAGEN Consortium, NNIPPS Study Group, I.B., G.A.N., D.B.R., R.P., M.C.K., J.G., O.W.W., T.R., B.S., I.K., C.A.H., P.N.L., F.C., A. Chiò, E.B., E.P., R.T., G.L., J.P., A.C.L., J.H.W., W.R., P.V.D., L.F., T.P., R.H.B., J.D.G., J.E.L., O. Hardiman, P.M.A., P.C., P.V., V.S., M.A.v.E., A.A.-C., L.H.v.d.B. and J.H.V. were involved in phenotyping, sample collection and management. W.v.R., A.S., A.M.D., R.L.M., F.P.D., R.A.A.v.d.S., P.T.C.v.D., G.H.P.T., M.K., A.M.B., W.S., A.R.J., K.P.K., I.F., A.V., N.T., R.D.S., W.J.B., A.V., K.V., M.R.-G., B.K., L.L., S. Abdulla, K.S., E.P., F.P.D., J.M., C. Curtis, G. Breen, A.A.-C. and J.H.V. prepared DNA and performed SNP array hybridizations. W.v.R., S.L.P., K.P.K., K.L., A.M.D., P.T.C.v.D., G.H.P.T., K.R.v.E., P.I.W.d.B. and J.H.V. were involved in the next-generation sequencing analyses. W.v.R., K.R.v.E., A. Menelaou, P.I.W.d.B., A.A.-C. and J.H.V. performed the imputation. W.v.R., A.S., F.P.D., R.L.M., S.L.P., S.d.J., I.F., N.T., W.S., A.R.J., K.P.K., K.R.v.E., K.S., H.M.B., P.I.W.d.B., M.A.v.E., C.M.L., G. Breen, A.A.-C., L.H.v.d.B. and J.H.V. performed GWAS analyses. W.v.R., A.M.D., R.A.A.v.d.S., R.L.M., C.R.A., M.K., A.M.B., R.D.S., E.P.M., J.A.F., C. Tunca, H.H., K.Z., P.C., P.V. and J.H.V. performed the replication analyses. W.v.R., A.S., R.L.M., M.R.R., J.Y., N.R.W., P.M.V., C.M.L., A.A.-C. and J.H.V. performed polygenic risk scoring and heritability analyses. S.d.J., U.V., L.F., T.H.P., W.v.R., O. Harschnitz, G. Breen, R.J.P. and J.H.V. performed biological pathway analyses. U.V., L.F., W.v.R. and J.H.V. performed eQTL analyses. W.v.R., A.S., A.A.-C., L.H.v.d.B. and J.H.V. prepared the manuscript with contributions from all authors. A.A.-C., L.H.v.d.B. and J.H.V. directed the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hardiman, O., van den Berg, L.H. & Kiernan, M.C. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7**, 639–649 (2011).
2. Al-Chalabi, A. *et al.* An estimate of amyotrophic lateral sclerosis heritability using twin data. *J. Neurol. Neurosurg. Psychiatry* **81**, 1324–1326 (2010).
3. van Es, M.A. *et al.* Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41**, 1083–1087 (2009).
4. Laaksovirta, H. *et al.* Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol.* **9**, 978–985 (2010).
5. Shatunov, A. *et al.* Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol.* **9**, 986–994 (2010).
6. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9orf72* causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
7. Renton, A.E. *et al.* A hexanucleotide repeat expansion in *C9orf72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
8. Fogh, I. *et al.* A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **23**, 2220–2231 (2014).
9. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
10. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
11. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
12. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
13. Höglinger, G.U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* **43**, 699–705 (2011).
14. Irwin, D.J. *et al.* Myelin oligodendrocyte basic protein and prognosis in behavioral-variant frontotemporal dementia. *Neurology* **83**, 502–509 (2014).
15. Cirulli, E.T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441 (2015).
16. Freischmidt, A. *et al.* Haploinsufficiency of *TBK1* causes familial ALS and fronto-temporal dementia. *Nat. Neurosci.* **18**, 631–636 (2015).
17. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
18. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
19. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
20. Wray, N.R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
21. Johnston, C.A. *et al.* Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. *J. Neurol.* **253**, 1642–1643 (2006).
22. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
23. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
24. Ramakrishnan, N.A., Drescher, M.J. & Drescher, D.G. The SNARE complex in neuronal and sensory cells. *Mol. Cell. Neurosci.* **50**, 58–69 (2012).
25. Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M. & Shaw, P.J. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7**, 616–630 (2011).
26. Lai, C.K. *et al.* Functional characterization of putative cilia genes by high-content analysis. *Mol. Biol. Cell* **22**, 1104–1119 (2011).
27. Ma, X., Peterson, R. & Turnbull, J. Adenylyl cyclase type 3, a marker of primary cilia, is reduced in primary cell culture and in lumbar spinal cord *in situ* in G93A SOD1 mice. *BMC Neurosci.* **12**, 71 (2011).
28. Krohn, K. *et al.* Immunochemical characterization of a novel mitochondrially located protein encoded by a nuclear gene within the DFNB8/10 critical region on 21q22.3. *Biochem. Biophys. Res. Commun.* **238**, 806–810 (1997).
29. Fang, X. *et al.* The NEK1 interactor, C21orf2, is required for efficient DNA damage repair. *Acta Biochim. Biophys. Sin. (Shanghai)* **47**, 834–841 (2015).
30. Vériène, J., Fossouo, L. & Parker, J.A. Neurodegeneration in *C. elegans* models of ALS requires TIR-1/Sarm1 immune pathway activation in neurons. *Nat. Commun.* **6**, 7319 (2015).

Wouter van Rheenen[1,123], Aleksey Shatunov[2,123], Annelot M Dekker[1], Russell L McLaughlin[3], Frank P Diekstra[1], Sara L Pulit[4], Rick A A van der Spek[1], Urmo Võsa[5], Simone de Jong[6,7], Matthew R Robinson[8], Jian Yang[8], Isabella Fogh[2,9], Perry TC van Doormaal[1], Gijs H P Tazelaar[1], Max Koppers[1,10], Anna M Blokhuis[1,10], William Sproviero[2], Ashley R Jones[2], Kevin P Kenna[11], Kristel R van Eijk[1], Oliver Harschnitz[1,10], Raymond D Schellevis[1], William J Brands[1], Jelena Medic[1], Androniki Menelaou[4], Alice Vajda[12,13], Nicola Ticozzi[9,14], Kuang Lin[2], Boris Rogelj[15,16], Katarina Vrabec[17], Metka Ravnik-Glavač[17,18], Blaž Koritnik[19], Janez Zidar[19], Lea Leonardis[19], Leja Dolenc Grošelj[19], Stéphanie Millecamps[20], François Salachas[20–22], Vincent Meininger[23,24], Mamede de Carvalho[25,26], Susana Pinto[25,26], Jesus S Mora[27], Ricardo Rojas-García[28,29], Meraida Polak[30,31], Siddharthan Chandran[32,33], Shuna Colville[32], Robert Swingler[32], Karen E Morrison[34], Pamela J Shaw[35], John Hardy[36], Richard W Orrell[37], Alan Pittman[36,38], Katie Sidle[37], Pietro Fratta[39], Andrea Malaspina[40,41], Simon Topp[2], Susanne Petri[42], Susanne Abdulla[43], Carsten Drepper[44], Michael Sendtner[44], Thomas Meyer[45], Roel A Ophoff[46–48], Kim A Staats[48], Martina Wiedau-Pazos[49], Catherine Lomen-Hoerth[50], Vivianna M Van Deerlin[51], John Q Trojanowski[51], Lauren Elman[52], Leo McCluskey[52], A Nazli Basak[53], Ceren Tunca[53], Hamid Hamzeiy[53], Yesim Parman[54], Thomas Meitinger[55], Peter Lichtner[55], Milena Radivojkov-Blagojevic[55], Christian R Andres[56], Cindy Maurel[56], Gilbert Bensimon[57–59], Bernhard Landwehrmeyer[60], Alexis Brice[61–65], Christine A M Payan[57,59], Safaa Saker-Delye[66], Alexandra Dürr[67], Nicholas W Wood[68], Lukas Tittmann[69], Wolfgang Lieb[69], Andre Franke[70], Marcella Rietschel[71], Sven Cichon[72–76], Markus M Nöthen[72,73], Philippe Amouyel[77], Christophe Tzourio[78], Jean-François Dartigues[78], Andre G Uitterlinden[79,80], Fernando Rivadeneira[79,80], Karol Estrada[79], Albert Hofman[80,81], Charles Curtis[6,7], Hylke M Blauw[1], Anneke J van der Kooi[82], Marianne de Visser[82], An Goris[83], Markus Weber[84], Christopher E Shaw[2], Bradley N Smith[2], Orietta Pansarasa[85], Cristina Cereda[85], Roberto Del Bo[86], Giacomo P Comi[86], Sandra D'Alfonso[87], Cinzia Bertolin[88], Gianni Sorarù[88], Letizia Mazzini[89], Viviana Pensato[90], Cinzia Gellera[90], Cinzia Tiloca[9], Antonia Ratti[9,14], Andrea Calvo[91,92], Cristina Moglia[91,92], Maura Brunetti[91,92], Simona Arcuti[93], Rosa Capozzo[93], Chiara Zecca[93], Christian Lunetta[94], Silvana Penco[95], Nilo Riva[96], Alessandro Padovani[97], Massimiliano Filosto[97], Bernard Muller[98], Robbert Jan Stuit[98], PARALS Registry[99], SLALOM Group[99], SLAP Registry[99], FALS Sequencing Consortium[99], SLAGEN Consortium[99], NNIPPS Study Group[99], Ian Blair[100], Katharine Zhang[100], Emily P McCann[100], Jennifer A Fifita[100], Garth A Nicholson[100,101], Dominic B Rowe[100], Roger Pamphlett[102], Matthew C Kiernan[103], Julian Grosskreutz[104], Otto W Witte[104], Thomas Ringer[104], Tino Prell[104], Beatrice Stubendorff[104], Ingo Kurth[105], Christian A Hübner[105], P Nigel Leigh[106], Federico Casale[91], Adriano Chio[91,92], Ettore Beghi[107], Elisabetta Pupillo[107], Rosanna Tortelli[93], Giancarlo Logroscino[108,109], John Powell[2], Albert C Ludolph[60], Jochen H Weishaupt[60], Wim Robberecht[83,110,111], Philip Van Damme[83,110,111], Lude Franke[5], Tune H Pers[112–116], Robert H Brown[11], Jonathan D Glass[30,31], John E Landers[11], Orla Hardiman[12,13], Peter M Andersen[60,117], Philippe Corcia[56,118,119], Patrick Vourc'h[56], Vincenzo Silani[9,14], Naomi R Wray[8], Peter M Visscher[8,120], Paul I W de Bakker[4,121], Michael A van Es[1], R Jeroen Pasterkamp[10], Cathryn M Lewis[6,122], Gerome Breen[6,7], Ammar Al-Chalabi[2,124], Leonard H van den Berg[1,124] & Jan H Veldink[1,124]

[1]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands. [2]Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, King's College London, London, UK. [3]Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [4]Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. [5]Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands. [6]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [7]NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [8]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [9]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy. [10]Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands. [11]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [12]Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Ireland. [13]Department of Neurology, Beaumont Hospital, Dublin, Ireland. [14]Department of Pathophysiology and Tranplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milan, Italy. [15]Department of Biotechnology, Jožef Stefan Institute, Ljubljana, Slovenia. [16]Biomedical Research Institute BRIS, Ljubljana, Slovenia. [17]Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. [18]Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia. [19]Ljubljana ALS Centre, Institute of Clinical Neurophysiology, University Medical Centre Ljubljana, Ljubljana, Slovenia. [20]Institut du Cerveau et de la Moelle Epinière, INSERM U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Université Paris 06, UMRS 1127, Paris, France. [21]Centre de Référence Maladies Rares SLA IIe de France, Département de Neurologie, Hôpital de la Pitié-Salpêtrière, Paris, France. [22]GRC-UPMC SLA et Maladies du Motoneurone, Paris, France. [23]Ramsay Generale de Santé, Hôpital Peupliers, Paris, France. [24]Réseau SLA IIe de France, Paris, France. [25]Institute of Physiology, Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal. [26]Department of Neurosciences, Hospital de Santa Maria–CHLN, Lisbon, Portugal. [27]Department of Neurology, Hospital San Rafael, Madrid, Spain. [28]Neurology Department, Hospital de la Santa Creu i Sant Pau de Barcelona, Autonomous University of Barcelona, Barcelona, Spain. [29]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain. [30]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA. [31]Emory ALS Center, Emory University School of Medicine, Atlanta, Georgia, USA. [32]Euan MacDonald Centre for Motor Neuron Disease Research, Edinburgh, UK. [33]Centre for Neuroregeneration and Medical Research Council

Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK. [34]Faculty of Medicine, University of Southampton, Southampton, UK. [35]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK. [36]Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK. [37]Department of Clinical Neuroscience, Institute of Neurology, University College London, London, UK. [38]Reta Lila Weston Institute, Institute of Neurology, University College London, London, UK. [39]Sobell Department of Motor Neuroscience and Movement Disorders, Institute of Neurology, University College London, London, UK. [40]Centre for Neuroscience and Trauma, Blizard Institute, Queen Mary University of London, London, UK. [41]North-East London and Essex Regional Motor Neuron Disease Care Centre, London, UK. [42]Department of Neurology, Hannover Medical School, Hannover, Germany. [43]Department of Neurology, Otto von Güricke University Magdeburg, Magdeburg, Germany. [44]Institute of Clinical Neurobiology, University Hospital Würzburg, Würzburg, Germany. [45]Department of Neurology, Charité University Hospital, Humboldt University, Berlin, Germany. [46]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands. [47]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [48]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA. [49]Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. [50]Department of Neurology, University of California, San Francisco, San Francisco, California, USA. [51]Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [52]Department of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [53]Neurodegeneration Research Laboratory, Boğaziçi University, Istanbul, Turkey. [54]Neurology Department, Istanbul Medical School, Istanbul University, Istanbul, Turkey. [55]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. [56]INSERM U930, Université François Rabelais, Tours, France. [57]AP-HP, Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, Paris, France. [58]UPMC, Pharmacologie, Paris VI, Paris, France. [59]BESPIM, CHU de Nîmes, Nîmes, France. [60]Department of Neurology, Ulm University, Ulm, Germany. [61]INSERM U1127, Hôpital de la Pitié-Salpêtrière, Paris, France. [62]CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, Paris, France. [63]Sorbonne Universités, UPMC Paris 06, UMRS 1127, Hôpital de la Pitié-Salpêtrière, Paris, France. [64]Institut du Cerveau et de la Moelle Epinière, Hôpital de la Pitié-Salpêtrière, Paris, France. [65]AP-HP, Département de Génétique, Hôpital de la Pitié-Salpêtrière, Paris, France. [66]Genethon, CNRS UMR 8587, Evry, France. [67]Department of Medical Genetics, Institut du Cerveau et de la Moelle Epinière, Hôptial Pitié-Salpêtrière, Paris, France. [68]Department of Neurogenetics, Institute of Neurology, University College London, London, UK. [69]PopGen Biobank and Institute of Epidemiology, Christian Albrechts University Kiel, Kiel, Germany. [70]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. [71]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Faculty of Medicine Mannheim, University of Heidelberg, Heidelberg, Germany. [72]Institute of Human Genetics, University of Bonn, Bonn, Germany. [73]Department of Genomics, Life and Brain Center, Bonn, Germany. [74]Division of Medical Genetics, University Hospital Basel, University of Basel, Basel, Switzerland. [75]Department of Biomedicine, University of Basel, Basel, Switzerland. [76]Institute of Neuroscience and Medicine INM-1, Research Center Juelich, Juelich, Germany. [77]University of Lille, INSERM, CHU de Lille, Institut Pasteur de Lille, U1167-RID-AGE Risk Factor and Molecular Determinants of Aging Diseases, Lille, France. [78]Bordeaux University, ISPED, Centre INSERM U1219–Epidemiologie Biostatistique et CIC-1401, CHU de Bordeaux, Pôle de Santé Publique, Bordeaux, France. [79]Department of Internal Medicine, Genetics Laboratory, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands. [80]Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands. [81]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [82]Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. [83]Department of Neurosciences, Experimental Neurology, Leuven Research Institute for Neuroscience and Disease (LIND), KU Leuven–University of Leuven, Leuven, Belgium. [84]Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, St. Gallen, Switzerland. [85]Laboratory of Experimental Neurobiology, IRCCS 'C. Mondino' National Institute of Neurology Foundation, Pavia, Italy. [86]Neurologic Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [87]Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases, Università del Piemonte Orientale, Novara, Italy. [88]Department of Neurosciences, University of Padova, Padova, Italy. [89]Department of Neurology, Università del Piemonte Orientale, Novara, Italy. [90]Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. [91]'Rita Levi Montalcini' Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy. [92]Azienda Ospedaliera Città della Salute e della Scienza, Turin, Italy. [93]Department of Clinical Research in Neurology, University of Bari 'A. Moro' at Pia Fondazione 'Card. G. Panico', Tricase, Italy. [94]NEMO Clinical Center, Serena Onlus Foundation, Niguarda Ca' Granda Hostipal, Milan, Italy. [95]Medical Genetics Unit, Department of Laboratory Medicine, Niguarda Ca' Granda Hospital, Milan, Italy. [96]Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy. [97]Neurology Unit, Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy. [98]Project MinE Foundation, Rotterdam, the Netherlands. [99]A list of members appears in the **Supplementary Note**. [100]Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia. [101]University of Sydney, ANZAC Research Institute, Concord Hospital, Sydney, New South Wales, Australia. [102]Stacey MND Laboratory, Department of Pathology, University of Sydney, Sydney, New South Wales, Australia. [103]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [104]Hans Berger Department of Neurology, Jena University Hospital, Jena, Germany. [105]Institute of Human Genetics, Jena University Hospital, Jena, Germany. [106]Department of Neurology, Brighton and Sussex Medical School Trafford Centre for Biomedical Research, University of Sussex, Falmer, UK. [107]Laboratory of Neurological Diseases, Department of Neuroscience, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy. [108]Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari 'Aldo Moro', Bari, Italy. [109]Unit of Neurodegenerative Diseases, Department of Clinical Research in Neurology, University of Bari 'Aldo Moro' at Pia Fondazione Cardinale G. Panico, Tricase, Italy. [110]Vesalius Research Center, Laboratory of Neurobiology, VIB, Leuven, Belgium. [111]Department of Neurology, University Hospitals Leuven, Leuven, Belgium. [112]Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA. [113]Division of Genetics, Boston Children's Hospital, Boston, Massachusetts, USA. [114]Center for Basic Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA. [115]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [116]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [117]Department of Pharmacology and Clinical Neurosience, Umeå University, Umeå, Sweden. [118]Centre SLA, CHRU de Tours, Tours, France. [119]Federation des Centres SLA Tours and Limoges, LITORALS, Tours, France. [120]Diamantina Institute, University of Queensland Translational Research Institute, Brisbane, Queensland, Australia. [121]Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands. [122]Department of Medical and Molecular Genetics, King's College London, London, UK. [123]These authors contributed equally to this work. [124]These authors jointly directed this work. Correspondence should be addressed to A.A.-C. (ammar.al-chalabi@kcl.ac.uk) or J.H.V. (j.h.veldink@umcutrecht.nl).

ONLINE METHODS

The software packages used, their version, web source and references are described in **Supplementary Table 18**.

GWAS discovery phase and quality control. Details on the acquired genotype data from previously published GWAS are described in **Supplementary Table 1**. Methods for case and control ascertainment for each cohort are described in the **Supplementary Note**. All cases and controls gave written informed consent, and the relevant institutional review boards approved this study. To obtain genotype data for newly genotyped individuals, genomic DNA was hybridized to the Illumina OmniExpress array according to the manufacturer's protocol. Subsequent quality control included (i) removing low-quality SNPs and individuals from each cohort, (ii) combining unbalanced cohorts on the basis of nationality and genotyping platform to form case–control strata, (iii) removing low-quality SNPs, related individuals and population outliers per stratum and (iv) calculating genomic inflation factors per stratum. More details are described in the **Supplementary Note** and **Supplementary Figure 11**. The number of SNPs and individuals failing each quality control step per cohort and stratum is displayed in **Supplementary Tables 2–5**.

Whole-genome sequencing (custom reference panel). Individuals were whole-genome sequenced on the Illumina HiSeq 2500 platform using PCR-free library preparation and 100-bp paired-end sequencing, yielding a minimum of 35× coverage. Reads were aligned to the hg19 human genome build, and after variant calling (Isaac variant caller) additional SNV and sample quality control was performed (**Supplementary Fig. 12** and **Supplementary Note**). Individuals in our custom reference panel were also included in the GWAS in strata sNL2, sNL3 and sNL4.

Merging reference panels. All high-quality calls in the custom reference panel were phased using SHAPEIT2 software. After checking strand and allele inconsistencies, both the 1000 Genomes Project reference panel (release 05-21-2011)[31] and custom reference panel were imputed up to the union of their variants as described previously[32]. Variants with inconsistent allele frequencies between the two panels were removed.

Imputation accuracy performance. To compare the imputation accuracy between different reference panels, 109 unrelated ALS cases of Dutch ancestry sequenced by Complete Genomics and 67 ALS cases from the UK sequenced by Illumina were selected as a test panel. All variants not present on the Illumina Omni1M array were masked, and the SNVs on chromosome 20 were subsequently imputed back using four different reference panels (1000 Genomes Project, GoNL, custom panel and merged panel). Concordance between the imputed alleles and sequenced alleles was assessed in each allele frequency bin where allele frequencies were calculated from the Dutch samples included in the Genome of the Netherlands cohort.

GWAS imputation. Prephasing was performed for each stratum using SHAPEIT2 with the 1000 Genomes Project phase 1 (release 05-21-2011) haplotypes[31] as a reference panel. Subsequently, strata were imputed up to the merged reference panel in 5-Mb chunks using IMPUTE2. Imputed variants with a MAF <1% or INFO score <0.3 were excluded from further analysis. Variants with allele frequency differences between strata, defined as deviating by >10 s.d. from the normalized mean allele frequency difference between those strata and an absolute difference >5%, were excluded because they are likely to represent sequencing or genotyping artifacts. Imputation concordance scores for cases and controls were compared to assess biases in imputation accuracy (**Supplementary Table 19**).

Meta-analysis. Logistic regression was performed on imputed genotype dosages under an additive model using SNPTEST software. On the basis of scree plots, one to four principal components were included per stratum. These results were then combined in an inverse-variance-weighted, fixed-effect meta-analysis using METAL. No marked heterogeneity across strata was observed as the Cochrane's Q test statistics did not deviate from the null distribution ($\lambda = 0.96$). Therefore, no SNPs were removed owing to excessive heterogeneity. The genomic inflation factor was calculated, and the quantile–quantile plot is provided in **Supplementary Figure 3a**.

Linear mixed model. All strata were combined including SNPs that passed quality control in every stratum. Subsequently, genetic relationship matrices (GRMs) were calculated for each chromosome including all SNPs using the Genome-Wide Complex Trait Analysis (GCTA) software package. Each SNP was then tested in an LMM including a GRM composed of all chromosomes excluding the target chromosome (leave one chromosome out, LOCO). The genomic inflation factor was calculated, and the quantile–quantile plot is provided as **Supplementary Figure 3b**.

Replication. For the replication phase, independent ALS cases and controls from Australia, Belgium, France, Germany, Ireland, Italy, the Netherlands and Turkey that were not used in the discovery phase were included. A predesigned TaqMan genotyping assay was used to replicate rs75087725 and rs616147. Sanger sequencing was performed to replicate rs10139154 and rs7813314 (**Supplementary Table 20** and **Supplementary Note**). All genotypes were tested in a logistic regression per country and subsequently underwent meta-analysis.

Rare variant analysis in *C21orf2*. The burden of nonsynonymous rare variants in *C21orf2* was assessed in whole-genome sequencing data obtained from ALS cases and controls from the Netherlands, Belgium, Ireland, the UK and the United States. After quality control, the burden of nonsynonymous and loss-of-function mutations in *C21orf2* was tested for association in each country and meta-analysis was subsequently performed. More details are provided in the **Supplementary Note** and **Supplementary Figure 13**.

Polygenic risk scores. To assess the predictive accuracy of polygenic risk scores in an independent data set, SNP weights were assigned on the basis of the LMM (GCTA-LOCO) analysis in 18 of 27 strata. SNPs in high LD (r^2 >0.5) in a 250-kb window were clumped. Subsequently, polygenic risk scores for cases and controls in the nine independent strata were calculated on the basis of their genotype dosages using PLINK v1.9. To obtain the Nagelkerke r^2 and corresponding P values, these scores were then regressed on their true phenotype in a logistic regression where (on the basis of scree plots) the first three principal components, sex and stratum were included as covariates.

SNP-based heritability estimates. *GCTA-REML.* GRMs were calculated using GCTA software including genotype dosages passing quality control in all strata. On the basis of the diagonal of the GRM, individuals representing subpopulations that contained an abundance of rare alleles (diagonal values mean ±2 s.d.) were removed (**Supplementary Fig. 14a**). Pairs where relatedness (off-diagonal) exceeded 0.05 were removed as well (**Supplementary Fig. 14b**). The eigenvectors for the first ten principal components were included as fixed effects to account for more subtle population structure. The prevalence of ALS was defined as the lifetime morbid risk for ALS (that is, 1 in 400)[21]. To estimate the SNP-based heritability for all non-genome-wide-significant SNPs, the genotypes for the SNPs reaching genome-wide significance were modeled as fixed effects. The variance explained by the GRM therefore reflects the SNP-based heritability of all non-genome-wide-significant SNPs. SNP-based heritability partitioned by chromosome or MAF was calculated by including multiple GRMs, calculated on SNPs from each chromosome or in the respective frequency bin, in one model.

Haseman–Elston regression. The phenotype correlation–genotype correlation (PCGC) regression software package was used to calculate heritability on the basis of the Haseman–Elston regression including the eigenvectors for the first ten principal components as covariates. The prevalence was again defined as the lifetime morbid risk (1 in 400).

LD score regression. Summary statistics from GCTA-LOCO and LD scores calculated from European individuals in 1000 Genomes Project were used for LD score regression. Associated SNPs ($P < 5 \times 10^{-8}$) and variants not in HapMap 3 were excluded. Considering adequate correction

for population structure and distant relatedness in the LMM, the intercept was constrained to 1.0 (ref. 12).

Biological pathway analysis (DEPICT). Functional interpretation of associated GWAS loci was carried out using DEPICT, using locus definition based on 1000 Genomes Project Phase 1 data. This method prioritizes genes in the affected loci and predicts involved pathways, biological processes and tissues, using gene co-regulation data from 77,840 expression arrays. Three separate analyses were performed for GWAS loci reaching $P = 1 \times 10^{-4}$, $P = 1 \times 10^{-5}$ or $P = 1 \times 10^{-6}$. One thousand permutations were used for adjusting the nominal enrichment *P* values for biases and additionally 200 permutations were used for FDR calculation.

31. Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
32. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).

doi:10.1038/ng.3622